INFORMATION AND CONTROL 17, 145–160 (1970)

Syntax and Semantics: A Categorical View

DAVID B. BENSON

Department of Computer and Information Science,
University of North Carolina, Chapel Hill, North Carolina 27514

A syntax is a category of strings and derivations between them. The semantic domain is a category of sets and functions. An interpretation is a cofunctor from the syntax to the semantics generated from a correspondence between productions and certain functions. There is a Galois connection between congruences on derivations and classes of interpretations. The smallest congruence of interest, similarity, is shown to correspond to the class of all interpretations. By considering certain subclasses of interpretations and the corresponding congruences, three different versions of "context-sensitive" are explicated.

1. INTRODUCTION

Thompson (1966) proposed a notion of semantics for formal languages generated by a semithue system in which the interpretation of a derivation is a function. Implications of essentially these semantic notions for context-free languages, in particular the implications for compiler design, were found by Knuth (1968). These semantic notions bear strong resemblance to the semantic methods for logical calculi (cf. Cohn, 1965). In the theory of models of logical calculi, the interpretations are to relational systems, and certain collections of interpretations play an important role.

In programming languages, and formal language theory the semantics is an action, or function, on certain sets. For example, the retrieval of information or the call and sequential processing of a subroutine is such an action. Notions of truth are not central here. The classes of interpretations are, however, useful.

Hotz (1966) (also see Schnorr, 1969) introduced the idea of using category theory to study the derivational structure imposed on a formal language by a rewriting system generating the language. The categorical formulation is refreshingly clear in its control of the details. A derivational structure, here called a syntax, is a certain type of category generated by a semithue system. The derivational richness of a syntax depends on the intent of the study, as

643/17/2-3

described in the next paragraph. The semantic domain is a category of sets and functions. The interpretations of a syntax are confunctors from the syntax to the semantic domain, generated by an association of functions with each production of the semithue system. The author has noted the apparent relatedness of Lawvere's "theories" (cf. Eilenberg and Wright, 1967).

In somewhat greater detail, the syntax category has as objects strings of letters drawn from some fixed alphabet and, if one chooses, only those strings derivable from the axiom of the semithue system. The morphisms are the derivations of one string from another. If the derivations are purely syntactic, or free, then it has been generally recognized for some time that inessential distinctions are made. One may say that the derivational structure is too rich. Hotz (1966) and Griffiths (1968) solved this by considering a certain relation between free derivations called similarty, which is shown to be an equivalence relation under more general conditions than are needed here. In fact, the relation is a congruence for the composition of derivations. The equivalence classes are again called derivations by Hotz. The semantic theory shows this is entirely suitable by relating derivational congruences and classes of interpretations in analogy to the methods in algebraic logic. The relation of similarity is shown to correspond to the class of all interpretations.

Even sparser derivational systems can be studied, by taking larger equivalence classes corresponding to smaller classes of interpretations. The application given here of the theory is the explication of three differing notions of context sensitive systems. Each variety is defined without semantics in the literature. We consider increasingly restricted subclasses of interpretations to obtain each of the three.

2. Notation and Terms

A category is a collection of objects and a collection of morphisms. Morphisms with domain a and codomain b are written $a \rightarrow b$ unless labeled: for example, $x : a \rightarrow b$. The set of morphisms from a to b is denoted by (a, b) There is at least one morphism in (a, a) for every object a, the identity on a. If $x \in (a, b)$ and $y \in (b, c)$ then the composition yx is a member of (a, c). Composition is associative. A functor from one category to another is a pair of functions: The object function maps objects to objects and the morphism function maps morphisms to morphisms, preserving identities and morphism composition. See MacLane and Birkhoff (1967), Mitchell (1965), Freyd (1964) or Cohn (1965) for exact definitions.

Lower case Greek letters denote strings over the alphabet A. The exceptions are ϕ, denoting the empty set, and $\in$, denoting class membership. The null string is denoted by λ. The set of all strings is A^*, the set of all nonnull strings is $A^+ = A^* - \{\lambda\}$. $N = \{0, 1, 2,...\}$, is the set of natural numbers.

A semithue system, as used here, is a triple $G = (A, P, \sigma)$ such that: A is an alphabet, P, the set of productions, is a finite binary relation in A^+ and σ, the axiom of G, is a member of A^+ (Davis, 1958, p. 84). The members of P are written $\alpha \to \beta$ to follow common usage and since we intend to consider them morphisms of a category.

The definition of derivation below is not the usual one, being equivalent to that in Griffiths (1968). The crucial fact is that derivations, as defined here, include sufficient information to completely specify the rewriting actions without ambiguity. In addition, the particular form of derivations used here possess the algebraic advantages that a subderivation is a derivation, that the natural composition of derivations is again a derivation, and that if $\theta \to \psi$ is a derivation, so is its extension by μ and ν, $\mu\theta\nu \to \mu\psi\nu$. Furthermore, the length zero derivations are naturally bijective with strings, giving the categorical identities, and the productions are naturally certain derivations of length one. The reader satisfied by the above description and uninterested in the details may skip the rest of this section.

A derivation from θ to ψ is a triple of finite sequences. The first member of the triple is a proof (Nelson, 1968, p. 88), i.e., the sequence of rewritten strings. In the formal language literature it has been common to denote by derivation this sequence alone. The second member of the triple is the sequence of productions applied in rewriting the strings. The third member is the sequence of left and right contexts in which the productions are applied.

DEFINITION 2.1. A derivation from θ to ψ is a triple

$$((\theta_0 ,..., \theta_n), (r_0 ,..., r_{n-1}), (\mu_0 - \nu_0 ,..., \mu_{n-1} - \nu_{n-1}))$$

for some $n \in N$, such that $\theta_0 = \theta$, $\theta_n = \psi$ and for all $i < n$, $\theta_i = \mu_i\alpha_i\nu_i$, $\theta_{i+1} = \mu_i\beta_i\nu_i$, where $r_i = \alpha_i \to \beta_i \in P$.

DEFINITION 2.2. If

$$x_1 = ((\theta_0 ,..., \theta_n), (r_0 ,..., r_{n-1}), (\mu_0 - \nu_0 ,..., \mu_{n-1} - \nu_{n-1}))$$

is a derivation from θ to ψ and

$$x_2 = ((\psi_0 ,..., \psi_m), (q_0 ,..., q_{m-1}), (\pi_0 - \rho_0 ,..., \pi_{m-1} - \rho_{m-1}))$$

is a derivation from ψ to ξ, then the composition of x_1 and x_2, denoted by x_2x_1, is the derivation (d, r, c) from θ to ξ such that

$$d = (\theta_0 ,..., \theta_n , \psi_1 ,..., \psi_m),\ r = (r_0 ,..., r_{n-1} , q_0 ,..., q_{m-1}),$$

and

$$c = (\mu_0 \_ \nu_0 ,..., \mu_{n-1} \_ \nu_{n-1} , \pi_0 \_ \rho_0 ,..., \pi_{m-1} \_ \rho_{m-1}).$$

DEFINITION 2.3. $x_2 = (d, r, c)$, a derivation from $\mu\theta\nu$ to $\mu\psi\nu$, is called the (μ, ν)-extension of $x_1 = ((\theta_0 ,..., \theta_n), q, (\pi_0 \_ \rho_0 ,..., \pi_{n-1} \_ \rho_{n-1}))$, a derivation from θ to ψ, iff $d = (\mu\theta_0\nu,..., \mu\theta_n\nu)$, $r = q$, and $c = (\mu\pi_0 \_ \rho_0\nu,..., \mu\pi_{n-1} \_ \rho_{n-1}\nu)$.

DEFINITION 2.4. The length zero derivation $((\theta), (\ ), (\ ))$ is called the θ-identity derivation.

DEFINITION 2.5. The length one derivation $((\alpha, \beta), (\alpha \to \beta), (\lambda \_ \lambda))$ is called the $(\alpha \to \beta)$ derivation.

3. SYNTAX

Three categories, called syntactic categories, are defined. Each of these categories demonstrates the free derivational structure G-induced on A^+. The complete category, **F**, is a G-induced relational algebra in categorical clothing. The objects of **F** are all the strings of A^+ and the set of morphisms (θ, ψ) is the set of derivations from θ to ψ. **F** is a category since (i) composition of derivations, which is associative, is the morphism composition, and (ii) for each object θ the θ-identity derivation is the categorical identity for θ under composition of derivations.

For each $\alpha \to \beta$ a production of P, the length one $(\alpha \to \beta)$ derivation has domain α, codomain β, and production application $\alpha \to \beta$. This morphism is purposely confused with the production $\alpha \to \beta$ so that the productions are morphisms of **F**.

For each $x : \theta \to \psi$, a morphism of **F**, and for each pair of objects μ, ν, there is a unique morphism of **F**, $\mu\theta\nu \to \mu\psi\nu$, which is the (μ, ν)-extension of x. In an extension of x, the only action is from θ to ψ with μ and ν remaining unchanged throughout the derivation $\mu\theta\nu \to \mu\psi\nu$.

Morphisms are called derivations when the linguistic structure is to be emphasized. The linguistic interest in **F** is as a convenient method of defining the following subcategory.

The syntax, **A**, is the full subcategory of **F** such that (θ, ψ) are morphisms of **A** just in case (σ, θ) is not empty in **F**. The objects of **A** are recovered from the identities remaining in **A**, and are the strings in A^* derivable from the axiom, σ.

If we divide the alphabet A into disjoint alphabets I and E, the internal and external alphabets respectively, we define the external syntax, **E**, as the full subcategory of **A** such that (θ, ψ) are morphisms of **E** just in case there is an $\omega \in E^+$ such that (ψ, ω) is not empty in **A**. The strings in E^+ which are also objects of **E** form the language of the grammar G, in the usual sense. The term "external" is used to avoid conflict with the use of the term "terminal" in category theory. The external syntax is not considered in the sequel.

The following facts are immediate. If G is monogenic then **A** is ordered. G is loop free if and only if for each object θ of **A**, (θ, θ) contains only the identity. If P is a binary relation from $A^+ - E^+$ to A^+, then any morphism of **A** with codomain a member of E^+ is epic. **A** does not, in general, have terminal objects.

4. Semantics and Interpretations

The semantic domain is taken to be some fixed category of sets and functions, **S**, in which cartesian products are available. Cartesian product is considered to be an associative operation. With this, the extension of functions is defined as follows: If W, X, Y, and Z are sets and $f : X \to Y$ any function from X to Y, then $t : W \times X \times Z \to W \times Y \times Z$ is an extension of f preserving W and Z iff t acts on Y as does f and acts as the identity on W and Z. That is, t is an extension of f if the following three diagrams commute, where the unlabeled arrows are the projections.

$$\begin{tikzcd} W \times X \times Z \arrow[r, "t"] \arrow[d] & W \times Y \times Z \arrow[d] \\ X \arrow[r, "f"] & Y \end{tikzcd}$$

$$\begin{tikzcd} W \times X \times Z \arrow[rr, "t"] \arrow[dr] & & W \times Y \times Z \arrow[dl] \\ & W & \end{tikzcd}$$

$$\begin{tikzcd} W \times X \times Z \arrow[rr, "t"] \arrow[dr] & & W \times Y \times Z \arrow[dl] \\ & Z & \end{tikzcd}$$

An interpretation of the syntax $\mathbf{A}$ is a cofunctor $I : \mathbf{A} \to \mathbf{S}$ taking strings to products and derivations to functions. The presentation is eased by considering interpretations on the complete category $\mathbf{F}$ followed by restriction to the subcategory $\mathbf{A}$.

DEFINITION 4.1. A cofunctor $I : \mathbf{F} \to \mathbf{S}$ is an interpretation iff (i) and (ii) hold.

(i) The object function of I satisfies $I(a) \neq \phi$ for each $a \in A$ and $I(\alpha a) = I(\alpha) \times I(a)$ for $\alpha \in A^+$ and $a \in A$.

(ii) The morphism function of I satisfies the property that for each derivation $\theta \to \psi$ of $\mathbf{F}$ and for each pair $\mu, \nu \in A^*$, the interpretation of the (μ, ν)-extension of $\theta \to \psi$, $\mu\theta\nu \to \mu\psi\nu$, is the function

$$I(\mu\theta\nu \to \mu\psi\nu) : I(\mu\psi\nu) \to I(\mu\theta\nu)$$

which is the extension of $I(\theta \to \psi) : I(\psi) \to I(\theta)$ preserving $I(\mu)$ and $I(\nu)$.

To specify a certain interpretation it suffices to give a function from the alphabet to sets, and a function, I, from the set of productions to functions such that if $\alpha \to \beta \in P$ then $I(\alpha \to \beta) : I(\beta) \to I(\alpha)$. If $I(\alpha)$ is a product of k factors and $I(\beta)$ is a product of n factors then $I(\alpha \to \beta)$ is a k-tuple of n-ary functions. If the cofunctor $I : \mathbf{F} \to \mathbf{S}$ is an interpretation then the restriction of I to $\mathbf{A}$, $I : \mathbf{A} \to \mathbf{S}$, is called an interpretation. The image of an interpretation is called the semantics of the interpretation. In general the semantics is not a subcategory of $\mathbf{S}$.

If $\omega \in E^+$ and $\sigma \to \omega$ is a derivation in $\mathbf{A}$, where σ is the axiom, then ω is a sentence. The interpretations of the sentence ω are functions from $I(\omega)$ to $I(\sigma)$. The meanings of ω are the images of the interpretations of the sentence ω. ω is ambiguous, in the usual syntactic sense, if it has more than one interpretation under I. The converse does not hold.

As an example, consider a syntax for addition of the natural numbers. The alphabet is unusual in order to avoid confusion between syntactic and semantic entities. The alphabet is {$, #, @}, the productions are

$$\$ \to \$@\$$$

$$\$ \to \#$$

and the axiom is $. The intent is that # denote the numeral one; that @ denote the addition sign; and that $ denote the syntactic class, expression. The usual symbols with these denotations are reserved to denote the corresponding semantic entities. The interpretation of the alphabet is given by

$I(\$) = N$, the set of natural numbers; $I(\#) = \{1\}$, the set whose sole member is the natural number 1; $I(@) = \Lambda$, a postulated identity for the cartesian product operation. One could avoid using Λ, but it simplifies the presentation of the example. The interpretation of the production set is given by

$$I(\$ \to \$@\$) : N^2 \to N = +: N^2 \to N,$$

the ordinary addition of natural numbers;

$$I(\$ \to \#) : \{1\} \to N,$$

the injection of 1 into the set of natural numbers.

There are two distinct phrase markers for derivations from \$ to $\#@\#@\#$, but each derivation has the interpretation $f: \{1\} \times \{1\} \times \{1\} \to N$ such that $f(1, 1, 1) = 3$. Exercising the terminology, the interpretation of the sentence $\#@\#@\#$ is $f: \{1\} \times \{1\} \times \{1\} \to N$; the meaning of $\#@\#@\#$ is the image of f, $\{3\}$; the semantics of the interpretation I is in this case a subcategory of **S**. The example suggests the resemblance of this formulation to that of tree automata (cf. Thatcher, 1967).

5. Congruences

Let $\sim$ be an equivalence relation defined on each set (θ, ψ) of morphisms. If two morphisms do not share the same domain and the same codomain then they are inequivalent. Suppose the equivalence, $\sim$, is such that for all $\eta, \theta, \psi, \omega, w \in (\eta, \theta)$, $x, y \in (\theta, \psi)$, $z \in (\psi, \omega)$ it is the case that $x \sim y$ implies $xw \sim yw$ and $zx \sim zy$. Then $\sim$ is called a congruence. The equivalence classes into which each (θ, ψ) is partitioned by the congruence $\sim$ are the morphisms of the quotient category $\mathbf{A}/\sim$ (Mitchell, 1965, p. 4).

Let Φ be the class of interpretations of **A**. For each congruence $\sim$ define $\Xi(\sim)$ as the class of all $I \in \Phi$ such that whenever $x \sim y$, then $I(x) = I(y)$. For each subclass Ξ define similarity modulo Ξ, $\sim(\Xi)$, by : $x \sim y(\Xi)$ iff x and y share domains and share codomains and for all $I \in \Xi$, $I(x) = I(y)$. The correspondences

$$\sim \to \Xi(\sim)$$

$$\Xi \to \sim(\Xi)$$

form a Galois connection (cf. Cohn, 1965, p. 44) between congruence relations and classes of interpretations, the pertinent facts being recorded in the following proposition.

PROPOSITION 5.1. *Let $\sim_1$ and $\sim_2$ be congruences on* **A**, *Ξ_1 and Ξ_2 subclasses of Φ. Then*

$$\sim_1 \subseteq \sim_2 \quad \textit{implies} \quad \Xi(\sim_1) \supseteq \Xi(\sim_2),$$

$$\Xi_1 \subseteq \Xi_2 \quad \textit{implies} \quad \sim(\Xi_1) \supseteq \sim(\Xi_2),$$

$$\sim_1 \subseteq \sim(\Xi(\sim_1)),$$

$$\Xi_1 \subseteq \Xi(\sim(\Xi_1)). \quad \blacksquare$$

Intuitively, two derivations are congruent if they possess the same structure, therefore imposing the same collection of possible meanings on the derived sentence. The Galois connection results in a semantic definition of structure. The structural description of a derivation is the equivalence class to which it belongs, and by the Galois connection this is a function of the particular subclass of interpretations considered acceptable by some external criteria.

The uninterpreted notion of structural description arises from the congruence relation of similarity, defined by Griffiths (1968). Let $\sim$ denote the similarity relation for the rest of the paper. In this section we show that $\sim = \sim(\Phi)$. That is, two derivations are similar if and only if they are identically interpreted by each possible interpretation. This result can be considered an additional reason for studying the X-categories of G. Hotz. The free X-categories are the quotient categories of complete syntaxes modulo similarity. If $\Xi(\approx)$ is a proper subclass of Φ for some congruence $\approx$, the corresponding quotient category is a nonfree X-category.

To begin, the definition of uninterpreted, or syntactic, similarity is required. Similarity is the least congruence relation on **A** such that if

$$x : \theta \to \psi, \; y : \theta' \to \psi'$$

are derivations of **A** with extensions

$$x_0 : \theta\theta' \to \psi\theta',$$

$$x_1 : \theta\psi' \to \psi\psi',$$

$$y_0 : \theta\theta' \to \theta\psi',$$

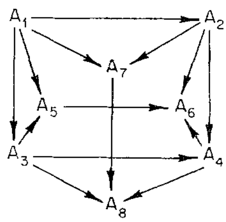
$$y_1 : \psi\theta' \to \psi\psi',$$

and $u = y_1 x_0$, $v = x_1 y_0$, then u is similar to v, $u \sim v$.

To motivate this definition, a brief description of Griffiths' development is given. A derivation can be interchanged into another by switching a pair of production applications when the applications do not interact. From the earlier exposition, $x_1 y_0$ and $y_0 x_1$ can be interchanged. A derivation can be L-interchanged if it can be interchanged and the resulting derivation acts earlier on some prefix of the domain. Thus $x_1 y_0$ may be L-interchanged. There is exactly one canonical derivation in each similarity class. In the case of context-free systems, canonical derivations are isomorphic to leftmost derivations (Ginsburg, 1966, p. 30), and the similarity classes, $[\theta \to \psi]$, are isomorphic to the phrase markers of derivations of ψ from θ.

To show that $u \sim v$ implies $u \sim v(\Phi)$ the following categorical fact is useful.

PROPOSITION 5.2. *In the dual triangular prism*



if every square and triangle commutes except possibly the square with verticies A_1, A_2, A_3, A_4, and $\{A_4 \to A_6, A_4 \to A_8\}$ is a product, then the remaining square commutes.

Proof. Let f_{ij} denote the morphism $A_i \to A_j$.
Then

$$\begin{aligned} f_{56} f_{15} &= f_{56} f_{35} f_{13} \\ &= f_{46} f_{34} f_{13} \\ &= f_{26} f_{12} \\ &= f_{46} f_{24} f_{12} . \end{aligned}$$

Similarly $f_{78} f_{17} = f_{48} f_{34} f_{13} = f_{48} f_{24} f_{12}$. Since $\{f_{46}, f_{48}\}$ is a product, the morphisms $f_{56} f_{15}$ and $f_{78} f_{17}$ factor uniquely through it, thus $f_{34} f_{13} = f_{24} f_{12}$. ∎

PROPOSITION 5.3. *$u \sim v$ implies $u \sim v(\Phi)$.*

Proof. Assume $u \sim v$. If $u = y_1 x_0$ and $v = x_1 y_0$ where x_0, x_1, y_0, y_1 are extensions of x and y as described above, then $I(u) = I(x_0)\, I(y_1)$, $I(v) = I(y_0)\, I(x_1)$, and $I(x_0)$, $I(x_1)$ are extensions of $I(x)$ while $I(y_0)$, $I(y_1)$ are extensions of $I(y)$. We obtain the dual triangular prism with $I(\psi\psi') = A_1$, $I(\theta\psi') = A_2$, $I(\psi\theta') = A_3$, $I(\theta\theta') = A_4$, $I(\psi) = A_5$, $I(\theta) = A_6$, $I(\psi') = A_7$, and $I(\theta') = A_8$. Every square and triangle commutes by the properties of extensions and the above categorical fact. In particular, $I(u) = I(v)$. Since the generating relation for similarity is a subset of $\sim(\Phi)$, it follows from the definition of similarity as the least congruence than $\sim \subseteq \sim(\Phi)$. ∎

To show that $u \sim v(\Phi)$ implies $u \sim v$ requires developing an interpretation, Q, which recovers derivations up to similarity. Since $Q \in \Phi$, the desired result will follow from showing that a canonical derivation similar to u can be obtained from $Q(u)$. Q may be thought of as the best Gödel numbering of derivations it is possible to obtain as an interpretation. The Gödel numbering is in the following generalized arithmetic, $G(A, P)$, over the alphabet A and productions P.

(i) $A \subseteq G(A, P)$.

(ii) $a \in A$, $r \in P$ and $g \in G(A, P)$ implies $(a, r, g) \in G(A, P)$.

(iii) Finite sequences of members of $G(A, P)$ are members of $G(A, P)$.

The required projections from $G(A, P)$ are denoted as follows:

$$\text{If } g = (a, r, g') \in G(A, P) \text{ then } p_0 g = a,\ p_1 g = r, \text{ and } p_2 g = g'.$$

Define the interpretation Q on letters of the alphabet by, for $a \in A$, $Q(a) = \{a\} \cup \{(a, r, g) \mid r \in P \,\&\, g \in G(A, P)\}$. Q is defined on the production $r = \alpha \to \beta$ where $\alpha = a_1 \cdots a_k$, as $Q(r) : Q(\beta) \to Q(\alpha)$ such that for each $g \in Q(\beta)$, $Q(r)(g) = ((a_1, r, g), \ldots, (a_k, r, g))$. Q is then extended to $\mathbf{A}$ as in Section 4.

If u is a derivation from θ to ψ define the Gödel number of u, $Gd(u)$, by $Gd(u) = Q(u)(\psi)$. Given $Gd(u) = (g_1, \ldots, g_m)$ the domain of u is recoverable as $a_1 \cdots a_m$ where $a_i = g_i$ if g_i is not a triple and $a_i = p_0 g_i$ if g_i is a triple. Recovery of the codomain of u from $Gd(u)$ is implicit in the recovery of the canonical derivation similar to u.

Proposition 5.4. *u and v are similar iff they have the same Gödel number.*

Proof. $u \sim v$ implies $Q(u) = Q(v)$, thus $Gd(u) = Gd(v)$. Assume that $Gd(u) = Gd(v)$. It suffices to assume, in addition, that v is canonical. Application of the following procedure recovers v from $Gd(u)$ in a finite number of iterations.

Assume $Gd(u) = (g_1, \ldots, g_m)$, $g_i \in G(A, P)$. For the largest i such that each of $g_1, \ldots, g_i$ is a letter and therefore not a triple, no production was applied to the prefix $\gamma = g_1 \cdots g_i$. If $i = m$ the construction is complete. If $i \neq m$ consider $g_{i+1} = (a, r, g')$ where $a \in A, r \in P, g' \in G(A, P)$. If $r = \alpha \to \beta$

the length of α is k, α is equal to the concatenate of $p_0 g_{i+1}, \ldots, p_0 g_{i+k}$, each of $p_1 g_{i+1}, \ldots, p_1 g_{i+k}$ is equal to r, and $g' = p_2 g_{i+1} = \cdots = p_2 g_{i+k}$, then the production $r = \alpha \rightarrow \beta$ is applied at this point in the canonical derivation v. The next finite sequence to which the reduction process applies is constructed by replacing, in $(g_1, \ldots, g_m)$, the subsequence $(g_{i+1}, \ldots, g_{i+k})$ by the sequence g'. The procedure is iterated on the new sequence.

In the above test for the applicability of $r = \alpha \rightarrow \beta$ if there is some j, $1 \leqslant j \leqslant k$, such that $\alpha_j \neq p_0 g_{i+j}$ or $r \neq p_1 g_{i+j}$ or $g' \neq p_2 g_{i+j}$, the production r must be defered since some other production application occured first, overlapping the application under test at g_{i+j}. In this case begin testing at position $i + j$.

u is similar to some canonical derivation, say x. Hence $Q(u) = Q(x)$ and $Gd(u) = Gd(x) = Gd(v)$. Since x is canonical by the above construction it is v. Thus $u \sim v$. ▮

From Propositions 5.3 and 5.4 we have

PROPOSITION 5.5. *For u, v derivations of* $\mathbf{A}$, *$u \sim v$ iff $u \sim v(\Phi)$.* ▮

As corollaries note that if Ω is a class of interpretations with $Q \in \Omega$ then $\sim(\Omega) = \sim$ and $\Xi(\sim(\Omega)) = \Phi$. The congruence of equality on $\mathbf{A}$ has as its corresponding class of interpretations Φ, thus $\sim(\Xi(=)) = \sim$.

6. Context Sensitive Systems

There are at least three different notions of the structural descriptions of sentences generated by a context-sensitive grammar. The first is the purely syntactic view in which the idea of allowing the rewriting by a context-free production only in a fixed (and local) context is avoided by requiring only that the length of production antecedents be less than or equal to the length of the corresponding production consequents. In this sense, the semithue production $\alpha \rightarrow \beta$ is type one context sensitive just in case the length of α is less than or equal to the length of β. It is well-known that the above is weakly equivalent to the more stringent condition on productions. However the corresponding structural descriptions lack the intuitively desired strength. Let P be a set of type one context sensitive productions and (A, P, s) generate the syntax $\mathbf{C}$. The class of interpretations of $\mathbf{C}$ is denoted by Φ and the structural descriptions of derivations in $\mathbf{C}$ are the morphisms of $\mathbf{C}/\sim$. In the purely syntactic view no additional structure is imposed. $\mathbf{C}$ is called a type one context sensitive syntax.

The second and third notions arise from considering the context sensitive productions to be context-free productions applicable only in a specified local context. The productions are often written in the form,

$$a \to \beta/\gamma \_ \delta,$$

where a is an internal latter, $\beta \in A^+$, and γ, $\delta \in A^*$. γ and δ are the left and right contexts, respectively, restricting the applicability of $a \to \beta$. Such productions can be considered to be semithue productions, and the associated syntax type two or type three context-sensitive depending upon the subclass of interpretations. Type three corresponds to the notion that the underlying p-marker is the structural description. With a larger class of interpretations one obtains the structural descriptions of Kuno (1967), called here type two.

A context-sensitive production $a \to \beta/\gamma \_ \delta$ is a semithue production $\gamma a\delta \to \gamma\beta\delta$ for which the allowable interpretations, I, are to functions constant to $I(\gamma)$ and $I(\delta)$ in the product

$$I(\gamma a \delta) = I(\gamma) \times I(a) \times I(\delta).$$

The action of the function to $I(a)$ distinguishes type two from type three. If the structural descriptions are to be p-markers, then the interpretation of $\gamma a\delta \to \gamma\beta\delta$ must depend solely on an interpretation of $a \to \beta$. That is, the function $I(\gamma a\delta \to \gamma\beta\delta) : I(\gamma\beta\delta) \to I(\gamma a\delta)$ must be constant to $I(\gamma)$ and $I(\delta)$ but act as some $f : I(\beta) \to I(a)$ in the codomain $I(\gamma a\delta)$. In type two, the underlying function may take its arguments from all of $I(\gamma\beta\delta)$, $f : I(\gamma\beta\delta) \to I(a)$.

Formally, a type two context-sensitive syntax is a type one context-sensitive syntax, **C**, such that with each production x, $x \in (\gamma a\delta, \gamma\beta\delta)$, of the semithue system generating **C**, is associated its context $\gamma \_ \delta$. This situation is denoted by $x = a \to \beta/\gamma \_ \delta$. Note that the set $(\gamma a\delta, \gamma\beta\delta)$ may contain more than one production; such multiplicities are distinguished by their contexts. In addition the class of admissible interpretations, Ξ, is the largest subclass of Φ such that for each production $x = a \to \beta/\gamma \_ \delta$ and for each $I \in \Xi$, the following diagrams commute, where the unlabeled arrows are the projections.

$$\begin{tikzcd} I(\gamma\beta\delta) \arrow[rr, "I(x)"] \arrow[dr] & & I(\gamma a\delta) \arrow[dl] \\ & I(\gamma) & \end{tikzcd}$$

$$\begin{tikzcd} I(\gamma\beta\delta) \arrow[rr, "I(x)"] \arrow[dr] & & I(\gamma a\delta) \arrow[dl] \\ & I(\delta) & \end{tikzcd}$$

A type three context-sensitive syntax is a type two context-sensitive syntax, **C**, with class of admissible interpretations, Ω, the largest subclass of Ξ such that for each $I \in \Omega$ and each production $x = a \to \beta/\gamma \_ \delta$ generating **C** there exists a function

$$f(x, I) : I(\beta) \to I(a).$$

Ω and the functions $f(x, I)$ must satisfy the properties:

(i) The commutivity of

$$\begin{array}{ccc} I(\gamma\beta\delta) & \xrightarrow{I(x)} & I(\gamma a\delta) \\ \downarrow & & \downarrow \\ I(\beta) & \xrightarrow{f(x,I)} & I(a) \end{array}$$

where the projections are again unlabeled.

(ii) If $x = a \to \beta/\gamma \_ \delta$ and $y = a \to \beta/\pi \_ \rho$ are productions then for each $I \in \Omega, f(x, I) : I(\beta) \to I(a) = f(y, I) : I(\beta) \to I(a)$.

The latter property corresponds to considering $a \to \beta/\gamma \_ \delta$ and $a \to \beta/\pi \_ \rho$ to be a single context-free production $a \to \beta$ restricted to the context $\gamma \_ \delta$ or to $\pi \_ \rho$. Without this property the underlying p-marker cannot be obtained. The structural description of a derivation x in **C** is the $\sim(\Omega)$ equivalence class $[x]_\Omega$. To each equivalence class corresponds a unique p-marker of the underlying context-free syntax, as the following demonstrates.

Suppose P is the set of productions which together with their contexts generate the type three context-sensitive syntax **C** with class of interpretations Ω. Then $R = \{a \to \beta \mid a \to \beta/\gamma \_ \delta \in P\}$ is the set of context-free productions generating the context-free syntax, **D**, with p-markers $\mathbf{D}/\sim$. The function $F : P \to R$ with $F(a \to \beta/\gamma \_ \delta) = a \to \beta$ extends to the context-freeing functor $F : \mathbf{C} \to \mathbf{D}$. For each $I : \mathbf{C} \to \mathbf{S}$ in Ω define $J_I : \mathbf{D} \to \mathbf{S}$ by

(i) For each object θ, $J_I(\theta) = I(\theta)$

(ii) On morphisms by induction from the productions R, where for each $a \to \beta \in R$, $J_I(a \to \beta) = f(x, I)$ for any $x = a \to \beta/\gamma \_ \delta \in P$ such that $F(x) = a \to \beta$.

The diagram

$$\begin{tikzcd} \mathbf{C} \arrow[rr, "F"] \arrow[dr, "I"'] & & \mathbf{D} \arrow[dl, "J_I"] \\ & \mathbf{S} & \end{tikzcd}$$

commutes for each $I \in \Omega$. Furthermore $\{J_I \mid I \in \Omega\}$ is equal to $\Phi_{\mathbf{D}}$, the entire class of interpretations of $\mathbf{D}$, since any $K \in \Phi_{\mathbf{D}}$ is extendible to some $I \in \Omega$.

From the above association of a context-free system with the type three context-sensitive syntax $\mathbf{C}$, one notes that the p-marker of x, a derivation in $\mathbf{C}$, is the similarity class of $F(x)$, $[F(x)]$.

PROPOSITION 6.1.

$$\begin{array}{ccc} \mathbf{C} & \xrightarrow{F} & \mathbf{D} \\ \downarrow & & \downarrow \\ \mathbf{C}/\Omega & \xrightarrow{G} & \mathbf{D}/\sim \end{array}$$

commutes where $F : \mathbf{C} \to \mathbf{D}$ *is the context-freeing functor for* $\mathbf{C}$, $\mathbf{C}/\Omega$ *is the quotient category of* $\mathbf{C}$ *modulo* $\sim(\Omega)$, $\mathbf{D}/\sim$ *is the quotient category of* $\mathbf{D}$ *modulo similarity, and* $G : \mathbf{C}/\Omega \to \mathbf{D}/\sim$ *is the structural description or p-marker functor, the identity on objects and on morphisms,* $G([x]_\Omega) = [F(x)]$.

Proof. G is well-defined since for any $x, y \in (\theta, \psi)$, if $x \sim y(\Omega)$ then for any $I \in \Omega$, $J_I(F(x)) = I(x) = I(y) = J_I(F(y))$, and so $F(x) \sim F(y)$. The commutivity of the diagram is obtained directly from the definition of G. ▮

PROPOSITION 6.2. *G is faithful, i.e., injective on morphisms.*

Proof. Suppose $G([x]_\Omega) = G([y]_\Omega)$, or equivalently, $F(x) \sim F(y)$. Since for all $K \in \Phi_{\mathbf{D}}$, $K(F(x)) = K(F(y))$, and since every $I \in \Omega$ is an extension of some $K \in \Phi_{\mathbf{D}}$, it is the case that for all $I \in \Omega$, $I(x) = I(y)$, so that $x \sim y(\Omega)$. ▮

From the propositions above one has

PROPOSITION 6.3. *If $x, y \in (\theta, \psi)$ are derivations of $\mathbf{C}$ then x and y are similar modulo Ω if and only if they have the same p-marker. That is, $x \sim y(\Omega)$ iff $F(x) \sim F(y)$.* ▮

Kuno (1967) defines trees augmented by quadruples of integers at each node as the structural descriptions of context sensitive derivations. The corresponding uninterpreted congruence relation here is called *cs*-similarity.

Cs-similarity, $\simeq$, is defined to be the least congruence on $\mathbf{C}$ such that

(i) $x \sim y$ implies $x \simeq y$, and

(ii) if $x : a \rightarrow \beta/\gamma \_ \mu\nu$ and $y : c \rightarrow \delta/\nu\pi \_ \rho$ are productions generating **C** with extensions

$$x_0 : \gamma a\mu\nu\pi c\rho \rightarrow \gamma\beta\mu\nu\pi c\rho$$

$$x_1 : \gamma a\mu\nu\pi\delta\rho \rightarrow \gamma\beta\mu\nu\pi\delta\rho$$

$$y_0 : \gamma a\mu\nu\pi c\rho \rightarrow \gamma a\mu\nu\pi\delta\rho$$

$$y_1 : \gamma\beta\mu\nu\pi c\rho \rightarrow \gamma\beta\mu\nu\pi\delta\rho$$

then $y_1 x_0 \simeq x_1 y_0$.

Two *cs*-similar derivations may be interchanged although the production applications overlap so long as the overlap is in the contexts and not in the underlying context free productions. Not surprisingly, *cs*-similarity is similarity modulo Ξ, where Ξ is the type two class of interpretations.

PROPOSITION 6.4. $\simeq \; = \; \sim(\Xi)$.

Proof. The proof is essentially the same as that of Proposition 5.5, but noting that the Gödel numbering interpretation Q is not in Ξ. An analogous Gödel numbering, Q', satisfying the type two restriction on interpretations is constructed to complete the proof. The definition of Q' on letters is identical to that of Q: For $a \in A, Q'(a) = \{a\} \cup \{(a, r, g) | \, r \in P \;\&\; g \in G(A, P)\}$. On productions, the situation differs. Let $r = a \rightarrow \beta/\gamma \_ \delta$ be a production where the length of γ is l, the length of β is m, and the length of δ is n. Then $Q'(r)$ is the function from $Q'(\gamma\beta\delta)$ to $Q'(\gamma a\delta)$ such that for $g \in Q'(\gamma\beta\delta)$, $g = (g_1, ..., g_l, b_1, ..., b_m, d_1, ..., d_n)$, $Q'(r)(g) = (g_1, ..., g_l, (a, r, g), d_1, ..., d_n)$. ▮

The type two notions are readily generalized to productions of the form $\alpha \rightarrow \beta/\gamma \_ \delta$ where $\alpha, \beta \in A^+$, $\gamma, \delta \in A^*$, and to pattern matching productions of the form

$$\alpha_0, \alpha_1, ..., \alpha_{n-1} \rightarrow \beta_0, \beta_1, ..., \beta_{n-1}/\gamma_0 \_ \gamma_1 \_ \cdots \gamma_{n-1} \_ \gamma_n,$$

which reduces to the previous form in the case $n = 1$.

ACKNOWLEDGMENTS

The author gratefully acknowledges the careful reading and helpful commentary of the referee, and support by the National Science Foundation.

RECEIVED: February 25, 1969; REVISED: March 30, 1970

REFERENCES

COHN, P. M. (1965), "Universal Algebra," Harper and Row, New York.

DAVIS, M. (1958), "Computability and Unsolvability," McGraw-Hill Book Co., New York.

EILENBERG, S. AND J. B. WRIGHT (1967), Automata in general algebras, *Information and Control* **11**, 452–470.

FREYD, P., (1964), "Abelian Categories," Harper and Row, New York.

GINSBURG, S. (1966), "The Mathematical Theory of Context-Free Languages," McGraw-Hill Book Co., New York.

GRIFFITHS, T. V. (1968), Some remarks on derivations in general rewriting systems, *Information and Control* **12**, 27–54.

HOTZ, G. (1966), Eindeutigkeit und Mehrdeutigkeit formaler Sprachen, *EIK*. **2**, 235–247.

KNUTH, D. E. (1968), Semantics of context-free languages, *Math. Systems Theory* **2**, 127–146.

MACLANE, S. AND G. BIRKHOFF (1967), "Algebra," Macmillian, New York.

MITCHELL, B. (1965), "Theory of Categories," Academic Press, Inc., New York.

KUNO, S. (1967), A context-sensitive recognition procedure, *in* Math. Linguistics and Auto. Translation Rpt. NSF-18, Aiken Comp. Lab., Harvard, Cambridge, Mass.

NELSON, R. J. (1968), "Introduction to Automata," John Wiley and Sons, Inc., New York.

SCHNORR, C. P. (1969), Transformational classes of grammars, *Information and Control* **14**, 252–277.

THATCHER, J. W. (1967), Characterizing derivation trees of context-free grammars through a generalization of finite automata theory, *J. Comp. System Sci.* **1**, 317–323.

THOMPSON, F. B. (1966), English for the computer, AFIPS Conf. Proc. Vol. **29**, 349–356, FJCC, Spartan Books, Washington, D. C.